NETIMIS: Dynamic Simulation of Health Economics Outcomes Using Big Data

Citation for published version:
Johnson, OA, Hall, PS & Hulme, C 2016, 'NETIMIS: Dynamic Simulation of Health Economics Outcomes Using Big Data', *PharmacoEconomics*, vol. 34, no. 2, pp. 107-114. https://doi.org/10.1007/s40273-016-0384-1

Digital Object Identifier (DOI):
10.1007/s40273-016-0384-1

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
PharmacoEconomics

Themed Issue on Big Data and its Application in Health Economics and Outcomes Research.

Theme - V. Emerging Methods of Analysis.

Leading Article.

NETIMIS: Dynamic Simulation of Health Economics Outcomes using Big Data

Authors: Owen A. Johnson[1, 2], Peter Hall[3], Claire Hulme[3]
Author Affiliations
1 – School of Computing, University of Leeds, Leeds, UK.
2 – X-Lab Ltd, Joseph's Well, Hanover Walk, Leeds, UK.
3 – Academic Unit of Health Economics, School of Medicine, University of Leeds, Leeds, UK.
Corresponding author: Owen A. Johnson o.a.johnson@leeds.ac.uk +44 113 3435474

Abstract
Many healthcare organizations are now making good use of e-health record (EHR) systems to record clinical information about their patients and the details of their healthcare. Electronic data in EHRs is generated by people engaged in complex processes within complex environments and their human input, albeit shaped by computer systems, is compromised by many human factors. This data is potentially valuable to health economists and outcomes researchers but is sufficiently large and complex enough to be considered part of the new frontier of "big data".

This paper describes emerging methods that draw together data mining, process modeling, activity based costing and dynamic simulation models. Our research infrastructure includes safe links to Leeds hospital's EHR with 3 million secondary and tertiary care patients. We created a multi-disciplinary team of health economists, clinical specialists, data and computer scientists and developed a dynamic simulation tool called NETIMIS (www.netimis.com) suitable for visualization of both human-designed and data-mined processes which can then be used for "what-if" analysis by stakeholders interested in costing, designing and evaluating healthcare interventions. We present two examples of model development to illustrate how dynamic simulation can be informed by big data from an EHR. We found the tool provided a focal point for multi-disciplinary team work to help them iteratively and collaboratively "deep dive" into big data.

Key Points for Decision Makers

- Big data can be defined as a data at a scale where research become uncomfortable – the point where existing methods are unable to unlock the value in the data.
- Electronic health records (EHR) are a type of big data that will be valuable to future health economists and outcomes researchers but the data is messy, it is the human product of large numbers of individuals working in complex environments. Solutions have yet to emerge.
- We found that the use of a dynamic simulation tool provided an effective focal point for multidisciplinary investigations into the messy world of EHR big data and the development of valuable insights.

1. Introduction

Many healthcare organizations are now making good use of e-health record (EHR) systems to record clinical information about their patients and track the care that they provide [1–4]. This data is sufficiently large and complex enough to be part of the new frontier of "big data", the subject of much recent literature [5–8], including in medicine and health [9–11]. "Big data" can be seen as data that is available at such a scale that traditional research and analysis methods fail. Researchers at this frontier are working collaboratively to discover new methods and approaches that will unlock the value of this new resource for research. In health, the big data in EHRs has the potential to be re-used to reduce the cost and transform the nature of research [12]. There are major challenges however - EHRs are still relatively new, their adoption has frequently been problematic [13], there are issues associated with research access to sensitive and confidential patient data [14] and there are concerns about the quality of this data [15] and therefore its suitability for research [16].

In this paper we describe our work in the UK with big data from a very large EHR. Our project brought together health economists, clinical specialists, data and computer scientists to combine data mining, process modeling, activity-based costing, and dynamic simulation. We developed a new dynamic simulation tool called NETIMIS and used it as a focal point for discussions on data quality and to help us understand existing care pathways. The NETIMIS tool can be used for hypothesis generation and exploratory research to support studies that evaluate healthcare interventions. Two recent publications from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force on Simulation Modeling Applications in Health Care Delivery Research [17,18] make a case that standardized methods for evaluating healthcare interventions (including decision trees and Markov modeling) are not sufficient for analyzing healthcare delivery systems. They advocate the use of dynamic simulation tools to help better understand the complex relationships in healthcare systems.

Our paper illustrates the challenges of working with EHR big data, the value of a collaborative approach and the interplay between a dynamic simulation tool and big data. To set the context we explore definitions of big data, the use of EHRs for research and the specific challenges around data quality. Our method is based on a unique research infrastructure, the NETIMIS tool and an iterative approach to model building. Two examples of model building are presented; the first illustrates the use of mixed methods to create a NETIMIS simulation of pathways in sepsis and the second shows how EHR data mining was used to automatically create a NETIMIS simulation in cancer care. In both cases collaborative working was the key to overcoming data quality issues.

2. Context: Big Data and Electronic Health Records

Big data as a term has its roots in the computing industry where increasing processing speed, storage capability and algorithmic techniques continually redefine the benchmark as to what is "normal" and what is "big". Definitions of big data tend to use the metrics of Volume, Velocity and Variety proposed by Laney [19], the 3V model, with other definitions adding additional Vs such as Value and Veracity. De Mauro et al. [20] propose a "consensual definition" of big data as "information assets characterized by such a High Volume, Velocity and Variety as to require specific technological and analytical methods for its transformation into Value", page 8. We would suggest the word "novel" rather than "specific" but the thrust of the definition is that "big" will remain relative to the methods available.

Electronic health records are an attractive source of big data for researchers. EHR systems evolved from paper-based physician notes and the requirement to structure these more formally, and eventually computerize, as health organizations have grown in size and complexity. Patient level information including demographic data and some clinical information (for example allergies, long term conditions) is supplemented by time-stamped records recording observations, diagnosis, prescriptions, treatment and administrative processes such as admission and discharge [21]. These events may be supplemented by attached images, documents and data files (for example ultrasound images, scans of letters received and biometric data) [22]. Event data will generally be a mixture of coded variables and natural language text logged against the date, time, user-id and type of event. There are therefore five elements available for big data analysis – patient level data, coded event data, natural language text, attached data files and the machine generated log files of events. Patient level data is of direct interest to population health. Coded event data may, or may not, follow national or international standards depending on the design and implementation of the system and wide variety in coding adherence is one of the challenges for analysis [11]. The event logs will contain longitudinal data which can be explored using process mining techniques [23–25]. In the UK, there has been mandatory use of clinical coding standards for several decades which makes UK EHR data a particularly rich source for data mining [26]. An alternative approach is to use Natural Language Processing (NLP) [27–29] of the unstructured notes and this has been particularly important in countries such as the USA where coding standards have not been widely adopted.

Even with widespread use of coding standards the variable nature of data quality in EHRs remains a major challenge for big data researchers. From a socio-technical perspective EHR data can be seen as the product of the continuously changing interplay between people, processes and informatics technology [30,31] . It follows that the provenance of EHR data matters when re-using the data for research [32]. Weiskopf and Weng [16] reviewed 95 journal papers which discussed EHR data quality and their thematic analysis identified five common quality dimensions (completeness, correctness, concordance, plausibility and currency) and seven quality assessment methods including comparison between data sources and "gold standards" determined through reviews with patients or clinicians.

3. Method

3.1 Aim and hypothesis

Our aim is to create accurate models of care provision from the evidence that can be mined from EHRs and then use these models as the basis for understanding the broader impact of interventions on healthcare systems. Our hypothesis is that issues with poor quality EHR data can be addressed iteratively by using a dynamic simulation tool as the focal point for sessions examining the data's provenance and context. Such an approach suggests a *deep dive* into the data, examining data from multiple perspectives and drawing on multiple sources.

3.2 Big Data Research Infrastructure

Our research is supported by the Leeds MRC Bioinformatics Research Centre which has established ethical access routes to PPM, the core EHR for Leeds Teaching Hospitals Trust (LTHT), an organization with six hospitals, 2,500 inpatient beds and 14,000 staff. The PPM EHR contains essentially episodic (episode of care) records for 3 million patients [33,34]. Initial work data mining cancer care pathways in PPM surfaced concerns with the veracity of data but also fascinating insights into patterns of care and their potential link with health outcomes [35]. Traces of care processes evident in the PPM event logs correlate to observations of the same process in the field and expert opinion drawn from literature and clinical advisers. There were however sufficient differences to indicate major challenges with all three sources – the logs reported treatment of cancer patients who had never been admitted, field observations noted variance in practice by practitioner and time of day while our clinical advisers were genuinely surprised to be shown what happened in practice.

3.3 The NETIMIS Dynamic Simulation Tool

The ISPOR Task Force on Simulation Modeling suggests dynamic simulation can help "enable a more realistic representation of the unique pathways of individual patients through the health care system" [18, page 10]. Our requirement was for a tool that could accurately model such pathways in a visually attractive form for interactive stakeholder participation, feedback and model refinement. The Task Force proposed a checklist (SIMULATE) to help determine when dynamic simulation is appropriate [17]. Following this checklist, our problem required modelling multiple events as processes (System), including non-linear relationships that make predicting outcomes difficult (Interactions), modelling systems at different levels (Multi-level), modeling complexity (Understanding), modeling feedback (Loops), interactions between entities (Agents), time-dependent behavior (Time) and surfacing emergent behavior (Emergence). The available tools can be broadly classified as based on systems dynamics, discrete event simulation or agent-based modelling though the Task Force note the increasing popularity of emerging hybrid models [18].

NETIMIS (Network Tools for Intervention Modelling with Intelligent Simulation) is classed as a discrete event simulation tool that includes some hybrid elements. It models individual entities as they flow through discrete events in a simulated process. The entities are not autonomous agents but have attributes that are randomized to reflect those of the base population and which can be used in decision rules as a proxy for agent based modelling [36]. It does not reproduce systems dynamic features such as stocks but does include others such as loops[17]. In common with both systems dynamics and agent-based modelling NETIMIS can model emergent properties such as non-linear behavior. The tool was developed to support our project by members of our team and a software company. Key design priorities were usability (the ability to quickly draw models and refine them without specialist expertise), visual impact (through color, design and animation), flexibility (to incorporate loops, queues, constraints, costs, times, and decision rules), accessibility (via Web browser) and performance (cloud-based processing).

Examples of NETIMIS pathways are shown in Figures 1, and 2. Care pathways are represented as a network of lines (with arrows to show direction) and nodes (box, diamond, parallelogram, dot etc. - used to show key activities and decision points). Colors, shapes, position, lengths and sizes are user configurable with most users adopting the convention from process modeling [37]. Simulation runs are animated by moving circles (tokens) representing individual patients that move through the network. Attributes of patient tokens are randomized to reflect those of the base population and pathway junctions are given probabilities that are dependent on those attributes. Patient colors can be configured by individual attribute values, for example a patient with two attributes (sex and age) may be colored blue (sex: female) and green (age band: 0-59) as in Fig 1. The user interacts with the software by drawing a care pathway using conventional drag and drop drawing tools and setting properties at model, population, line and node levels. Simulations can be run at any time and the user can adjust properties to create realistic and visually attractive representations of real-world pathways.

Outcomes are represented by multiple pathway end points and simulation runs calculate total costs and times based on the individual costs and times for each patient. The tool is pre-populated with reference sets of activity based costs [38] from published UK NHS sources (www.gov.uk/government/collections/nhs-reference-costs and www.pssru.ac.uk/project-pages/unit-costs) so that users can quickly create pathways that reflect current health economic costing models. Where such costs are not available estimated values can be used and verified through stakeholder engagement. By linking each of the activities in the model to reference costs, the simulation can be used to analyze the healthcare expenditures associated with care pathways.

3.4 Iterative Approach to Model Development

The model building strategy was based on agile methods [39] including time-boxed iterations and pair working (modeler and clinician). The team included health economists, software engineers, data miners, business analysts and clinicians from the study areas. The approach combined dynamic simulation modelling good practice [18], data mining methods (including statistical analysis, extract transform load, data cleansing), with modelling methods (care pathways [40], process modelling, economic modelling), primary data collection (including observation and interviews) [37] and data quality methods [16] (notably comparison between sources, validity checks and "gold standard" clinical review). The end of each iteration was marked by a team meeting to review the model with the focus moving from plausibility, to completeness to correctness with each session driving questions and suggested methods for the next iteration of investigation.

4. Example 1 - Dynamic Simulation of Care Pathways in Sepsis

This example was motivated by the need for a model to assess interventions using Point-of-care Testing (POCT) devices [41] that have the potential to improve the early detection of severe sepsis [42,43]. Model development was based on pair working. The modeler was an oncologist and health informatician and the clinician was Sepsis Lead for LTHT. The support team included the authors, a data miner and a second business analyst. Three iterations were completed and reviewed by the wider team. The resulting pathway simulation models are available online at www.netimis.com and are described more fully in [44]. The model for Community Acquired Pneumonia (CAP) is shown below (Fig. 1).

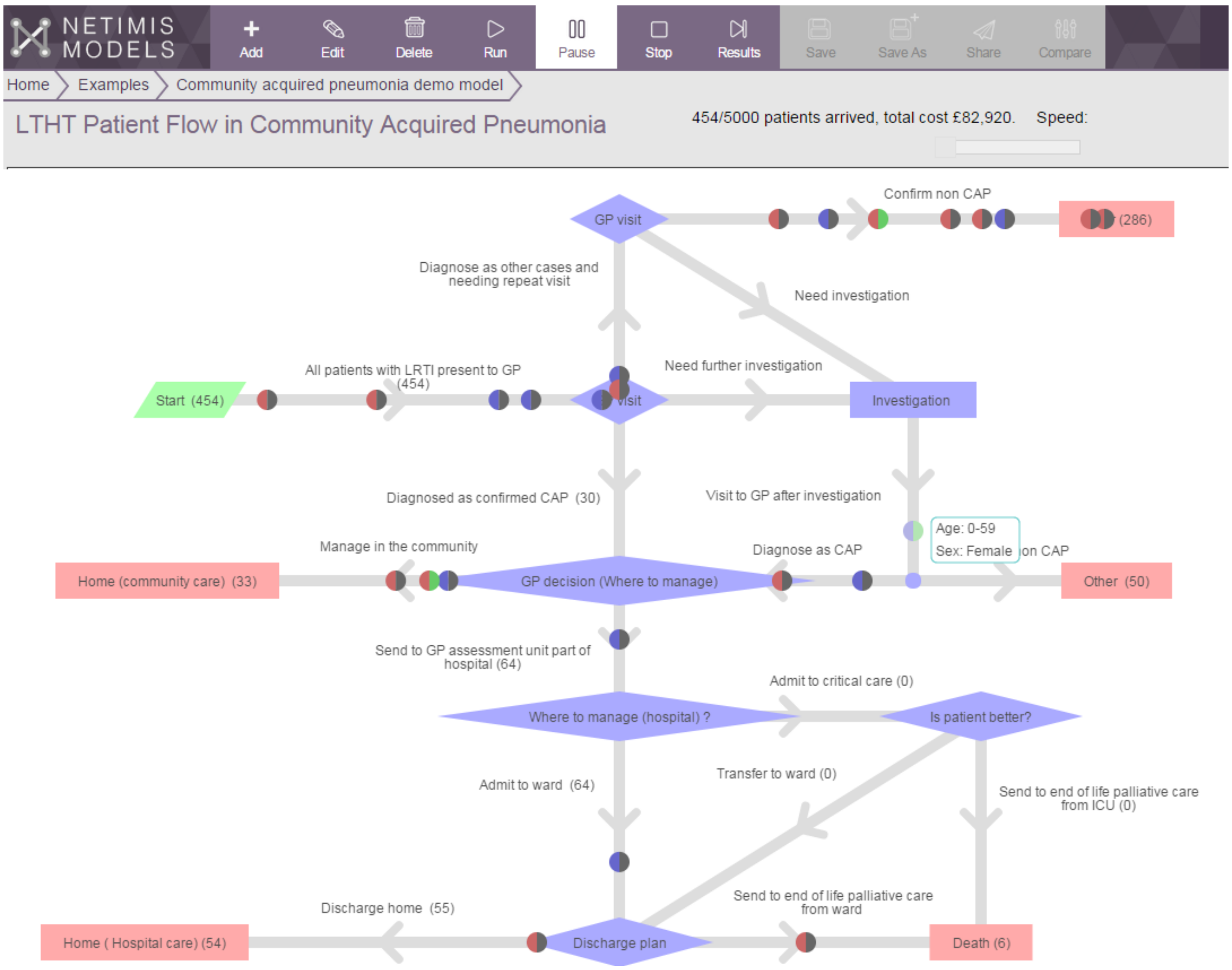

Fig. 1 NETIMIS Model of Community Acquired Pneumonia.

The probabilities for all decision points were based on literature sources for males and females over and under the age of 60 and the population randomized to reflect the split. Where data on probabilities and costs were not available (e.g. numbers transferred to hospital wards), assumptions were made with guidance from expert advice from the hospital. The results of modeling were compared to actual data mined from the PPM system and are shown below. Table 1 presents a comparison between two scenarios modelled with NETIMIS and an initial population of 1,000 patients presenting with lower respiratory tract infections (LRTI).

	CAP scenario 1		CAP scenario 2	
Path	Frequency	Total cost (GBP)	Frequency	Total cost (GBP)
Admit to critical care	5	6655	7	9317
Admit to ward	72	37800	65	34125
All patients with LRTI present to GP	1000	45000	1000	45000
Diagnosed as confirmed CAP	56	0	355	0
Discharge home	67	0	64	0
Send to GP assessment unit part of hospital	77	3465	72	3240
Send to end of life palliative care from ICU	2	0	2	0
Transfer to ward	3	1293	5	2155
TOTAL cost		157913		139122

Table 1 Results from two NETIMIS simulations of Community Acquired Pneumonia [44]

CAP Scenario 1 summarizes the outputs from the final validated model above and CAP scenario 2 models a "what if" scenario. In this case the early diagnosis of CAP by a hypothetical POCT device (note that the model can be adjusted to reflect varying measures of the efficacy of the device). In this comparison, the POCT device would save GBP 18,791 per 1,000 presenting patients and reduce adverse outcomes for patients. The break-even point for this scenario would therefore be GBP 18.79 per patient tested. The economic model could be refined to reflect the impact of adverse quality of life, cost to patients.

5. Example 2 - Dynamic Simulation of Chemotherapy Cycles

This example was motivated by the need to better understand how patterns of chemotherapy care have non-linear behavior over time. The modeler was a data miner and the clinician was Cancer Lead for LTHT. The support team included the authors, a business analyst and another oncologist. In this case event data was mined directly from the EHR but we found the initial quality issues too serious for meaningful results, a large number of both systemic (e.g. events with start times but no end times) and seemingly random errors (e.g. anomalous events) in the data were evident. Activities were renamed for readability and in some cases very similar activity names were combined (e.g. multiple names for many types of chemotherapy). We focused on the most frequent systemic errors first and, for each, tried to understand their root cause by audit and by discussing with clinicians engaged in the process to agree the most appropriate response. We then looked for patterns in errors to understand how they occurred and whether they had systemic causes. The result was an agreed *extract transform load* strategy, implemented in program code for the transformation of raw event data into a cleaned format. Eight iterations were completed and reviewed. An example of a partial run for 3,058 patients receiving EC90 for breast cancer is shown in Figure 2 (below).

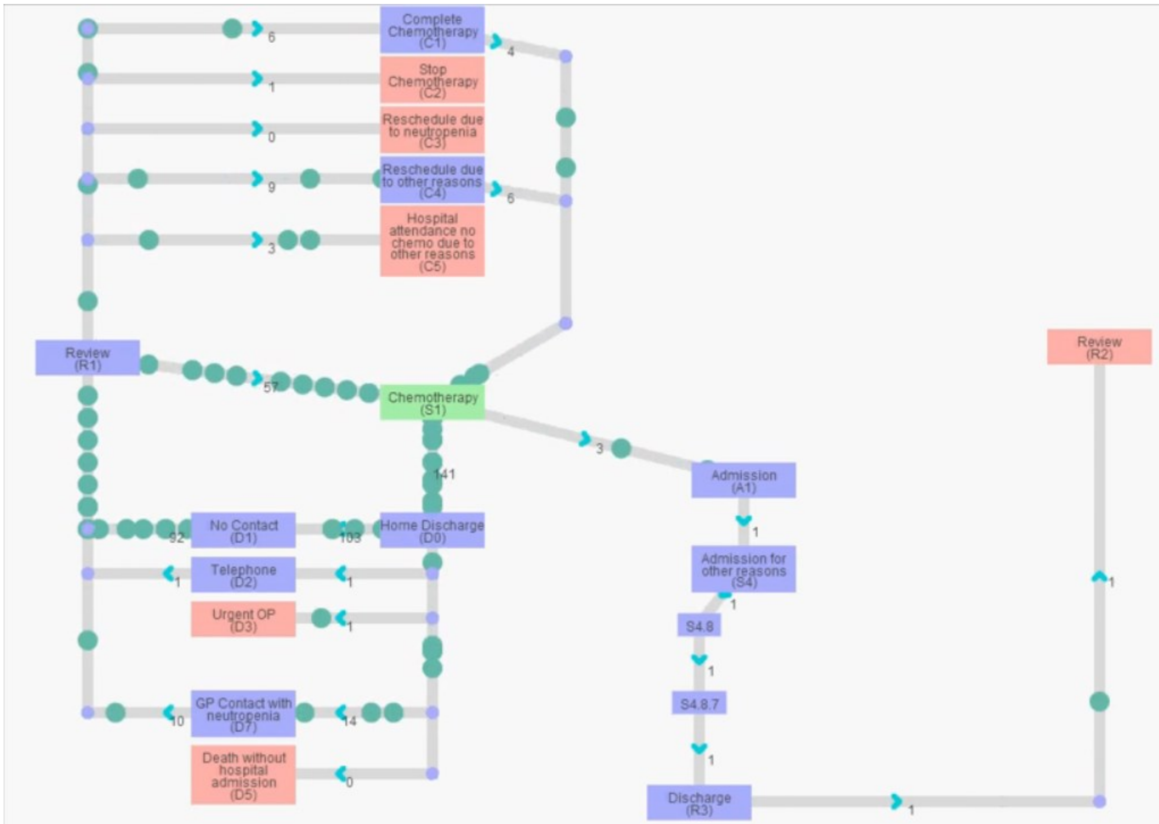

Fig. 2 NETIMIS model animating chemotherapy patient flow

In Figure 2 tokens representing real patients are shown as circles traversing pathways shown as grey lines. Note that the model is a visualization of actual events rather than a simulation, each circle represents a real patient from the EHR. The cyclical nature of chemotherapy is evident with the majority of patients cycling from Review to Chemotherapy to Discharge.

6. Summary

We found dynamic simulation to be effective for developing models that informed discussions about real-world patient pathways. Our use of a hybrid discrete event simulation tool and our approach to model building conformeds to the recommendations of the ISPOR Task Force and was well received by stakeholders. The NETIMIS tool was designed to support this engagement and proved to be effective. It is now available as a commercial product (www.netimis.com).

In both examples we were fortunate to have strong clinical support and access to other data sources. We have linked EHR data to dynamic simulation in two different ways. In Example 1, data mining provided some of the data to populate and validate our model but a mixture of other methods were needed to develop a complete model. In Example 2, the simulation is entirely based on data mined from the EHR but many secondary sources and considerable effort were required to address the data quality issues. We have demonstrated our hypothesis that this approach is possible but recognize that the process has been expensive and time consuming. One unexpected impact of working directly with clinical teams was their willingness to review, and indeed change, current practice as new insights were revealed. Similarly, feedback to the software engineers developing the EHR has helped identify changes which will improve future data quality.

The limitations of our work are that it has been highly contextual. We have used one dynamic simulation tool that was expressly built for our work so we can make no claims to generalizability. Although we have gained meaningful results from mining our big data EHR we had to employ other methods to address data quality issues. We were fortunate in having access to an EHR that is well established and well regarded in the UK, a country where clinical coding has been strong. With EHRs that are less well established and in countries with fewer incentives to follow coding standards it may not be possible to replicate the work for some years.

Acknowledgments

Owen Johnson is a Co-Investigator in the Leeds MRC Medical Bioinformatics Research Centre (PI: Professor Sir Alex Markham) which is funded by the UK Medical Research Council (Grant no. MR/L01629X/1). This Centre provided the research infrastructure within which the study took place. The authors have worked within a large multi-disciplinary group which includes multiple academic disciplines, central University of Leeds IT department, software developers from X-Lab and clinicians, managers and software developers from Leeds Teaching Hospital Trust. The work described in this paper is necessarily the output of a large number of these people.

Compliance with Ethical Standards

This study and the development of NETIMIS was funded by a UK Technology Strategy Board (now InnovateUK) Small Business Research Initiative grant (Grant no. TSB/971273) to X-Lab Ltd. OJ is a shareholder and director of X-Lab Ltd and was principle investigator for the grant. OJ and PH were employed to work on the study through their employer, the University of Leeds. CH acted in advisor to the study but received no remuneration. No external funding was received for the preparation of this article. Other than the connections declared above all three authors (OJ, PH, CH) have no conflicts of interest to declare. All authors contributed to the study design, interpretation of results and report writing.

REFERENCES

1. Adler-Milstein J, Everson J, Lee S-YD. Sequencing of EHR adoption among US hospitals and the impact of meaningful use. J. Am. Med. Inform. Assoc. [Internet]. 2014;21:984–91. Available from: http://dx.doi.org/10.1136/

2. Shu T, Liu H, Goss FR, Yang W, Zhou L, Bates DW, et al. EHR adoption across China's tertiary hospitals: A cross-sectional observational study. Int. J. Med. Inform. 2014;83:113–21.

3. Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. Br. Med. J. [Internet]. 2011;343. Available from: http://discovery.ucl.ac.uk/1350238/

4. Johnson OA, Fraser HS, Wyatt JC, Walley JD. Electronic Health Records in the UK and USA. Lancet. 2014;384:954.

5. Hey T. The next scientific revolution. Harv. Bus. Rev. 2010;88.

6. Graham-Rowe D, Waldrop M, Lynch C. Community cleverness required. Nature. 2008;455:1.

7. Constantiou ID, Kallinikos J. New games, new rules: big data and the changing context of strategy. J. Inf. Technol. [Internet]. 2015;30:44–57. Available from: http://www.palgrave-journals.com/doifinder/10.1057/jit.2014.17

8. Taylor L, Schroeder R, Meyer E. Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? Big Data Soc. [Internet]. 2014;1:1–10. Available from: http://bds.sagepub.com/lookup/doi/10.1177/2053951714536877

9. Huang T, Lan L, Fang X, An P, Min J, Wang F. Promises and Challenges of Big Data Computing in Health Sciences. Big Data Res. [Internet]. 2015;2:2–11. Available from: http://linkinghub.elsevier.com/retrieve/pii/S2214579615000118

10. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA [Internet]. 2013;309:1351–2. Available from: http://dx.doi.org/10.1001/jama.2013.393\nhttp://jama.jamanetwork.com/data/Journals/JAMA/926712/jvp130007_1351_1352.pdf

11. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Heal. Inf. Sci. Syst. [Internet]. 2014;2:3. Available from: http://www.hissjournal.com/content/2/1/3

12. Lobach DF, Detmer DE. Research Challenges for Electronic Health Records. Am J Prev Med Am. J. Prev. Med. 2007;32:104–11.

13. Boonstra A, Versluis A, Vos JFJ. Implementing electronic health records in hospitals: a systematic literature review. BMC Health Serv. Res. [Internet]. 2014;14:370. Available from: http://www.biomedcentral.com/1472-6963/14/370

14. Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. JAMA [Internet]. 2014;311:2479–80. Available from: http://jama.jamanetwork.com/article.aspx?articleid=1883026

15. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. Perspect. Heal. Inf. Manag. [Internet]. 2013;10:1c. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3797550&tool=pmcentrez&rendertype=abstract

16. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J. Am. Med. Informatics Assoc. [Internet]. 2013;20:144–51. Available from: http://jamia.oxfordjournals.org/cgi/doi/10.1136/amiajnl-2011-000681

17. Marshall DA, Burgos-Liz L, IJzerman MJ, Osgood ND, Padula W V., Higashi MK, et al. Applying Dynamic Simulation Modeling Methods in Health Care Delivery Research—The SIMULATE Checklist: Report of the ISPOR Simulation Modeling Emerging Good Practices Task Force. Value Heal. [Internet]. 2015;18:5–16. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1098301514047640

18. Marshall DA, Burgos-Liz L, Eng I, IJzerman MJ, Crown W, Padula W V, et al. Selecting a Dynamic Simulation Modeling Method for Health Care Delivery Research—Part 2: Report of the ISPOR Dynamic Simulation Modeling Emerging Good Practices Task Force.

19. Laney D. Application Delivery Strategies [Internet]. META Delta. 2001. Available from: http://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

20. De Mauro A, Greco M, Grimaldi M. What is Big Data? A Consensual Definition and a Review of Key Research Topics. AIP Conf. Proceedings, 1644 [Internet]. Madrid: AIP; 2015. p. 97–105. Available from: http://dx.doi.org/10.1063/1.4907823

21. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. Int. J. Med. Inform. 2008;77:291–304.

22. Coiera E. Guide to Health Informatics. 2nd ed. Hodder Arnold; 2003.

23. Van der Aalst W, Weijters T, Maruster L. Workflow Mining: Discovering process models from event logs. IEEE Trans. Knowl. Data Eng. 2004;16:1128–42.

24. Mans RS, Schonenberg MH, Song MS, van der Aalst WMP, Bakker PJM. Application of Process Mining in Healthcare : a Case Study in a Dutch Hospital. Proc. BIOSTEC 2008. Springer-Verlag, Berlin; 2009. p. 425–38.

25. Partington A, Wynn M, Suriadi S, Ouyang C, Karnon J. Process mining for clinical processes: a comparative analysis of four Australian hospitals. ACM Trans. Manag. Inf. Syst. [Internet]. 2015;9:1–18. Available from: http://eprints.qut.edu.au/66728/

26. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Ther Adv Drug Saf [Internet]. 2012;0:1–11. Available from: http://www.sagepub.co.uk/

27. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J. Am. Med. Inform. Assoc. 2011;18:544–51.

28. Hina S, Atwell E, Johnson OA. Semantic Tagging of Medical Narratives with Top Level Concepts from SNOMED CT Healthcare Data Standard. Int. J. Intell. Comput. Res. 2010;2:118–23.

29. Hina S, Atwell E, Johnson OA. Secure information extraction from clinical documents using SNOMED CT gazetteer and natural language processing. Internet Technol. Secur. Trans. (ICITST), Int. Conf. 2010. p. 1–5.

30. Whetton S. Health Informatics: A socio-technical perspective. Melbourne: Oxford University Press; 2005.

31. Berg M. Implementing information systems in health care organizations: myths and challenges. Int. J. Med. Inform. [Internet]. 2001;64:143–56. Available from: www.elsevier.com/locate/ijmedinf

32. Johnson KE, Kamineni A, Fuller S, Olmstead D, Wernli KJ. How the Provenance of Electronic Health Record Data Matters for Research: A Case Example Using System Mapping. eGEMs (Generating Evid. Methods to Improv. patient outcomes) [Internet]. 2014;2. Available from: http://repository.academyhealth.org/egems

33. Johnson OA, Abiodun SE. Understanding What Success In Health Information Systems Looks Like: The Patient Pathway Management (PPM) System At Leeds. UK Acad. Inf. Syst. Conf. Proc. [Internet]. 2011; Available from: http://aisel.aisnet.org/ukais2011/22

34. Newsham AC, Johnston C, Hall G, Leahy MG, Smith AB, Vikram A, et al. Development of an advanced database for clinical trials integrated with an electronic patient record system. Comput. Biol. Med. 2011;41:575–86.

35. Baker K, Goel R. Poster Presentation: Data mining to identify trends in response to chemotherapy. 3rd Int. Oncol. Eng. Conf. 2013.

36. Chhatwal J, He T. Economic Evaluations with Agent-Based Modelling: An Introduction. J PhamacoEconomics. 2015;33:423–33.

37. Holt J. A Pragmatic Guide to Business Process Modelling. British Computer Society; 2009.

38. Cooper R, Bray D, Parzen M. Who Wins in a Dynamic World : Theory of Constraints vs . Activity-Based Costing ? North Am. Assoc. Comput. Soc. Organ. Sci. Conf. June 2007 [Internet]. 2007; Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=962270

39. Dingsøyr T, Nerur S, Balijepally V, Moe NB. A decade of agile methodologies: Towards explaining agile software development. J. Syst. Softw. 2012;85:1213–21.

40. Crocker T, Johnson OA, King S. The suitability of care pathways for integrating processes and information systems in healthcare. Transform. Gov. People, Process Policy. 2009. p. 289–301.

41. Price CP, St John A, Krika LJ. Point-of-Care Testing: Needs, Opportunity, and Innovation. 3rd ed. AACC Press; 2010.

42. Minton J, Sandoe J, Wilcox M, McGann H, Nsutebu E, Gossiel A. Improving early management of severe sepsis. J. Infect. 2007;55:e70.

43. Tiru B, DiNino EK, Orenstein A, Mailloux PT, Pesaturo A, Gupta A, et al. The Economic and Humanistic Burden of Severe Sepsis. J PharmacoEconomics. 2015;33.

44. Mishra P, Clamp S, Johnson OA. NETIMIS : A healthcare pathway modelling simulation tool to view patient flow and impact of point of care testing [Internet]. 2014. Available from: https://www.netimis.co.uk/case-studies

Correspondence: *Owen A. Johnson*, School of Computing and Leeds MRC Bioinformatics Research Centre, The University of Leeds, Leeds, LS2 9JT, UK.

Email: o.a.johnson@leeds.ac.uk